

# Universal Gröbner Bases for Binary Linear Codes

Natalia Dück<sup>1</sup> and Karl-Heinz Zimmermann<sup>2</sup>

<sup>1,2</sup> Hamburg University of Technology  
21073 Hamburg, GERMANY

## Abstract

Each linear code can be described by a code ideal given as the sum of a toric ideal and a non-prime ideal. In this way, several concepts from the theory of toric ideals can be translated into the setting of code ideals. It will be shown that after adjusting some of these concepts, the same inclusion relationship between the set of circuits, the universal Gröbner basis and the Graver basis holds. Furthermore, in the case of binary linear codes, the universal Gröbner basis will consist of all binomials which correspond to codewords that satisfy the Singleton bound and a particular rank condition. This will give rise to a new class of binary linear codes denoted as Singleton codes.

**AMS Subject Classification:** 13P10, 94B05

**Key Words and Phrases:** Linear code, Gröbner basis, universal Gröbner basis, Graver basis, circuit, toric ideal, Singleton code

## 1 Introduction

When digital data are transmitted through a noisy channel errors can occur. But as receiving correct data is indispensable in many applications, error-correcting codes, which allow the detection and correction of a certain amount of errors by adding redundancy, are employed to tackle this problem. The construction of such codes and the study of their key properties is an ongoing and important task.

Gröbner bases, on the other hand, are a powerful tool that has originated from commutative algebra and provide a uniform approach to grasping a wide range of problems such as solving algebraic systems of equations, ideal membership decision, and effective computation in residue class rings modulo polynomial ideals [1, 2, 7, 15].

The first connection between linear codes and Gröbner bases was established in [6] which soon became known as the "Cooper philosophy". This link was based on the description of cyclic codes as ideals in a certain polynomial ring, where entries of a codeword are viewed as coefficients of a polynomial.

In [4], a different connection between linear codes and ideals in polynomial rings was presented, which was followed up in [13, 14]. In this approach, linear codes are described by a binomial ideal in a polynomial ring over an arbitrary field that can be written as the sum of a toric ideal and a non-prime ideal, the so-called code ideal. Toric ideals play a central role in this framework and are a well-studied special class of ideals in polynomial rings arising in various applications [3, 5, 8, 15]. This correspondence with linear codes proved to be extremely beneficial as it allowed the application of (slightly modified) results from the rich theory of toric ideals [14]. Furthermore, it has been shown that it allows an easy construction of the reduced Gröbner basis with respect to the lexicographic order from a systematic generator matrix [13]. Many results in algebraic geometry, however, depend on the chosen monomial order and thus knowledge of the universal Gröbner basis for this ideal, i.e., a finite generating set of the ideal that is a Gröbner basis for all monomial orders, is advantageous [17].

In this paper, some concepts used in connection with toric ideals are adapted to code ideals. It will be shown that for any code ideal the same inclusion relationship between the set of circuits, the universal Gröbner basis and the Graver basis holds as for toric ideals. In particular, for binary codes, all three sets will coincide. Furthermore, it will be proved that the universal Gröbner basis for a binary linear code has a neat structure: it consists of all binomials associated to codewords which satisfy the Singleton bound and a particular rank condition. This gives rise to a new class of codes denoted as Singleton codes.

This paper is organized as follows. The next section presents the basics about Gröbner bases and introduces toric ideals. In the third section, the main notions of linear codes are facilitated and their connection to ideals is established. The main results are contained in the fourth section.

## 2 Gröbner Bases and Toric Ideals

Throughout the paper, let  $\mathbb{K}$  denote a field,  $\mathbb{Z}$  the set of integers, and  $\mathbb{N}_0$  the set of non-negative integers. Denote by  $\mathbb{K}[\mathbf{x}] = \mathbb{K}[x_1, \dots, x_n]$  the polynomial ring in  $n$  indeterminates  $x_1, \dots, x_n$  and by  $\mathbf{x}^u = x_1^{u_1} x_2^{u_2} \dots x_n^{u_n}$  the *monomials* in  $\mathbb{K}[\mathbf{x}]$ , where  $u = (u_1, \dots, u_n) \in \mathbb{N}_0^n$ . The *total degree* of a monomial  $\mathbf{x}^u$  in  $\mathbb{K}[\mathbf{x}]$  is given by the sum of exponents  $|u| = u_1 + \dots + u_n$ . For any  $\omega \in \mathbb{R}^n$ , the  $\omega$ -degree of  $\mathbf{x}^u$  is defined by the usual inner product  $u \cdot \omega$  in  $\mathbb{R}^n$ . A *monomial order* on  $\mathbb{K}[\mathbf{x}]$  is a relation  $\succ$  on the set of monomials in  $\mathbb{K}[\mathbf{x}]$  satisfying: (1)  $\succ$  is a total ordering, (2) the monomial  $\mathbf{x}^0 = 1$  is the unique minimal element, and (3)  $\mathbf{x}^u \succ \mathbf{x}^v$  implies  $\mathbf{x}^u \mathbf{x}^w \succ \mathbf{x}^v \mathbf{x}^w$  for all  $u, v, w \in \mathbb{N}_0^n$ . Prominent monomial orders are the lexicographic order, the degree lexicographic order, and the degree reverse lexicographic order.

Given a monomial order  $\succ$ , each non-zero polynomial  $f \in \mathbb{K}[\mathbf{x}]$  has a unique

*leading term*, denoted by  $\text{lt}_{\succ}(f)$  or simply  $\text{lt}(f)$ , which is given by the largest involved term. The coefficient and the monomial of the leading term are called the *leading coefficient* and the *leading monomial*, respectively.

If  $I$  is an ideal in  $\mathbb{K}[\mathbf{x}]$  and  $\succ$  is a monomial order on  $\mathbb{K}[\mathbf{x}]$ , its *leading ideal* is the monomial ideal generated by the leading monomials of its elements,

$$\langle \text{lt}(I) \rangle = \langle \text{lt}(f) \mid f \in I \rangle. \quad (1)$$

A finite subset  $\mathcal{G}$  of an ideal  $I$  in  $\mathbb{K}[\mathbf{x}]$  is a *Gröbner basis* for  $I$  with respect to  $\succ$  if the leading ideal of  $I$  is generated by the set of leading monomials in  $\mathcal{G}$ ; that is,

$$\langle \text{lt}(I) \rangle = \langle \text{lt}(g) \mid g \in \mathcal{G} \rangle. \quad (2)$$

A Gröbner basis  $\mathcal{G}$  is *minimal* if no monomial in the set  $\mathcal{G}$  is redundant, and it is *reduced* if for any two distinct elements  $g, h \in \mathcal{G}$ , no term of  $h$  is divisible by  $\text{lt}(g)$  and all its elements have leading coefficient equal to 1. A reduced Gröbner basis for an ideal  $I$  with respect to a monomial order  $\succ$  is uniquely determined and will be denoted by  $\mathcal{G}_{\succ}(I)$ .

Gröbner bases solve the ideal membership problem. More concretely, a polynomial  $f$  belongs to an ideal  $I$  if and only if it is being reduced to zero on division by a Gröbner basis for  $I$ .

Gröbner bases can be computed by *Buchberger's algorithm* which is implemented in most computer algebra systems.

Although infinitely many monomial orders exist, there are only finitely many reduced Gröbner bases for a given ideal. The union of all reduced Gröbner bases is called the *universal Gröbner basis* for  $I$ . More details on Gröbner bases can be found in [1, 2, 7, 12].

Toric ideals form a specific class of ideals which can be defined in different ways [3]. For a subset  $A \subset \mathbb{Z}^d$  of  $n$  vectors or equivalently a matrix  $A \in \mathbb{Z}^{d \times n}$ , the *toric ideal*  $I_A$  is defined as

$$I_A = \langle \mathbf{x}^u - \mathbf{x}^v \mid Au = Av, u, v \in \mathbb{N}_0^n \rangle. \quad (3)$$

Each element  $u \in \mathbb{Z}^n$  can be uniquely written as  $u = u^+ - u^-$  where  $u^+, u^-$  have disjoint support and their entries are non-negative. Based on this, the toric ideal  $I_A$  can also be expressed as [15]

$$I_A = \left\langle \mathbf{x}^{u^+} - \mathbf{x}^{u^-} \mid u \in \ker(A) \right\rangle. \quad (4)$$

The binomials in the generating set (4) are *pure*, i.e., the greatest common divisor of the terms  $\mathbf{x}^{u^+}$  and  $\mathbf{x}^{u^-}$  in the binomial  $\mathbf{x}^{u^+} - \mathbf{x}^{u^-}$  is 1.

### 3 Linear Codes over Prime Fields

Let  $\mathbb{F}$  be a finite field and let  $n$  and  $k$  be positive integers with  $n \geq k$ . A *linear code* of length  $n$  and dimension  $k$  over  $\mathbb{F}$  is the image  $\mathcal{C}$  of a one-to-one linear mapping  $\phi : \mathbb{F}^k \rightarrow \mathbb{F}^n$ , i.e.,  $\mathcal{C} = \{\phi(a) \mid a \in \mathbb{F}^k\}$ . Such a code is denoted as  $[n, k]$  *code* and its elements are called *codewords*. In algebraic coding, the codewords are always written as row vectors. Alternatively, a code  $\mathcal{C}$  can be described as the row space of a matrix  $G \in \mathbb{F}^{k \times n}$ , whose rows form a basis of  $\mathcal{C}$ , and the matrix  $G$  is then called a *generator matrix* for  $\mathcal{C}$ . Any other generator matrix for  $\mathcal{C}$  can be obtained from a given generator matrix for  $\mathcal{C}$  by multiplying it from the left with a regular matrix. A code  $\mathcal{C}$  is in *standard form* if it has a generator matrix which is *systematic*, i.e.,  $G = (I_k \mid M)$ , where  $I_k$  is the  $k \times k$  identity matrix. Note that a generator matrix for an  $[n, k]$  code can contain a zero column. Such a code can be shortened by deleting this column giving an  $[n-1, k]$  code. All subsequently considered codes are assumed to have no such zero column.

Two  $[n, k]$  codes are *equivalent* if one can be obtained from the other by a monomial transformation, i.e., a linear map given by a monomial matrix, which is a matrix that has in each row and column exactly one non-zero element. It follows that every linear code is equivalent to a linear code in standard form.

The dual code  $\mathcal{C}^\perp$  of an  $[n, k]$  code  $\mathcal{C}$  over  $\mathbb{F}$  is an  $[n, n-k]$  code consisting of all words  $u \in \mathbb{F}^n$  such that  $u \cdot c = uc^T = 0$  for each  $c \in \mathcal{C}$ , where  $c^T$  denotes the transposed of  $c$ . If  $G = (I_k \mid M)$  is a generator matrix for  $\mathcal{C}$ , then  $H = (-M^T \mid I_{n-k})$  is a generator matrix for  $\mathcal{C}^\perp$ . For each word  $c \in \mathbb{F}^n$ , we have  $c \in \mathcal{C}$  if and only if  $Hc^T = \mathbf{0}$ . The matrix  $H$  is a *parity check matrix* for  $\mathcal{C}$ .

The *support* of a vector  $u \in \mathbb{F}^n$ , denoted by  $\text{supp}(u)$ , is the subset of  $\underline{n} = \{1, \dots, n\}$  given by all indices  $i \in \underline{n}$  with  $u_i \neq 0$ , and the *Hamming weight*, denoted by  $\text{wt}(u)$ , is the number of non-zero components and so equals the cardinality of the codeword's support. Note that for a binary code, each codeword is completely determined by its support. The *weight distribution* of an  $[n, k]$  code  $\mathcal{C}$  is a finite sequence of integers  $A_0, A_1, \dots, A_n$ , where  $A_i$ ,  $0 \leq i \leq n$ , denotes the number of codewords in  $\mathcal{C}$  having Hamming weight  $i$ . The *Hamming distance* between two vectors  $u, v \in \mathbb{F}^n$  is the number of positions in which they differ and so is given by the Hamming weight  $\text{wt}(u - v)$  of the difference vector. The Hamming distance defines a metric on  $\mathbb{F}^n$ . The minimum Hamming distance between any two distinct codewords in  $\mathcal{C}$  is the *minimum distance* of  $\mathcal{C}$ .

For any matrix  $G \in \mathbb{F}^{k \times n}$  and any subset  $J \subseteq \underline{n}$  of indices, let  $G_J$  denote the  $k \times |J|$  submatrix of  $G$  consisting of the columns with indices in  $J$ . Similarly, let  $c_J$  be the vector of length  $|J|$  consisting of the coordinates of  $c$  with indices in  $J$ . A subset  $J \subseteq \underline{n}$  of cardinality  $k$  is called an *information set* of the code if the  $k \times k$  submatrix  $G_J$  has rank  $k$ . In particular, the

following are equivalent:

1. The set of indices  $J$  is an information set.
2. For each  $m \in \mathbb{F}^k$  there is a unique  $c \in \mathcal{C}$  with  $c_J = m$ .
3. For every generator matrix  $G$  of the code  $\mathcal{C}$ ,  $G_J$  has rank  $k$ .

By the second assertion, a code cannot contain an information set  $J \subseteq \underline{n} \setminus \text{supp}(c)$  at the zero positions of a non-zero codeword  $c$ .

More basics on linear codes can be found in [10, 16].

For a given  $[n, k]$  code  $\mathcal{C}$  over a field  $\mathbb{F}_p$  with  $p$  elements, define the associated *code ideal* as

$$I_{\mathcal{C}} = \langle \mathbf{x}^c - \mathbf{x}^{c'} \mid c - c' \in \mathcal{C} \rangle + I_p, \quad (5)$$

where

$$I_p = \langle x_i^p - 1 \mid 1 \leq i \leq n \rangle. \quad (6)$$

Note that  $I_p$  allows to view the exponents of the monomials as vectors in  $\mathbb{F}_p^n$ . Each codeword  $c - c' \in \mathcal{C}$  with  $\text{supp}(c) \cap \text{supp}(c') = \emptyset$  can be associated with the binomial  $\mathbf{x}^c - \mathbf{x}^{c'}$  in the code ideal  $I_{\mathcal{C}}$ , where it is always assumed that  $\mathbf{x}^c \succ \mathbf{x}^{c'}$  whenever a monomial order  $\succ$  is considered. Note that unlike for toric ideals this binomial representation is not unique. Nevertheless, a binomial  $\mathbf{x}^c - \mathbf{x}^{c'}$  in  $I_{\mathcal{C}}$  is said to be associated with the codeword  $c - c'$ . Observe that the code ideal of a code  $\mathcal{C}$  can be based on a toric ideal as follows,

$$I_{\mathcal{C}} = I_A + I_p, \quad (7)$$

where  $A$  is an integral  $n - k \times n$  matrix such that  $H = A \otimes_{\mathbb{Z}} \mathbb{F}_p$  is a parity check matrix for  $\mathcal{C}$ .

## 4 Universal Gröbner Bases

In [13] it has been shown that the reduced Gröbner basis for the code ideal  $I_{\mathcal{C}}$  with respect to the lexicographic order can be read off from a generator matrix for the corresponding code  $\mathcal{C}$ .

**Theorem 1** ([13]). *Let  $\mathcal{C}$  be an  $[n, k]$  code over a prime field  $\mathbb{F}_p$  generated by a matrix  $G = (I_k \mid *)$  with row vectors  $g_i = e_i + m_i$ ,  $1 \leq i \leq k$ , where  $e_i$  denotes the  $i$ th unit vector and  $m_i$  a row vector of length  $n$  whose first  $k$  entries are zero. The reduced Gröbner basis with respect to any lexicographic order with  $\{x_1, \dots, x_k\} \succ \{x_{k+1}, \dots, x_n\}$  for the code ideal  $I_{\mathcal{C}}$  is given by*

$$\mathcal{G}_{\succ}(I_{\mathcal{C}}) = \{x_i - \mathbf{x}^{m_i} \mid 1 \leq i \leq k\} \cup \{x_i^p - 1 \mid k+1 \leq i \leq n\}.$$

This result can be further generalized.

**Corollary 2.** *Let  $\mathcal{C}$  be an  $[n, k]$  code over a prime field  $\mathbb{F}_p$  with an information set  $J = \{i_1, \dots, i_k\} \subseteq \underline{n}$ . There exists a generator matrix  $G$  in reduced row echelon form with respect to the columns indexed by  $J$  and row vectors  $g_{i_j} = e_{i_j} + m_j$ , where  $1 \leq j \leq k$ , and the reduced Gröbner basis with respect to any lexicographic order with  $\{x_j \mid j \in J\} \succ \{x_\ell \mid \ell \in \underline{n} \setminus J\}$  for the code ideal  $I_{\mathcal{C}}$  is given by*

$$\mathcal{G}_{\succ}(I_{\mathcal{C}}) = \{x_{i_j} - \mathbf{x}^{m_j} \mid i_j \in J\} \cup \{x_\ell^p - 1 \mid \ell \in \underline{n} \setminus J\}.$$

In [15] the author has introduced several concepts in the context of toric ideals which will be utilized in the following. However, since code ideals are not toric but a sum of a toric and a non-prime ideal, several of these concepts need to be adapted. In particular, it will become apparent that binary and non-binary codes need to be distinguished.

A binomial  $\mathbf{x}^c - \mathbf{x}^{c'}$  in  $I_{\mathcal{C}}$  is called *primitive* if there is no other binomial  $\mathbf{x}^u - \mathbf{x}^{u'}$  in  $I_{\mathcal{C}}$  such that  $\mathbf{x}^u$  divides  $\mathbf{x}^c$  and  $\mathbf{x}^{u'}$  divides  $\mathbf{x}^{c'}$ . Additionally, every binomial  $x_i^p - 1$  for  $1 \leq i \leq n$  is considered to be primitive. For binary codes, the case of  $c' = \mathbf{0}$  is excluded for reasons which will later become apparent. Note that if a monomial order  $\prec$  is given,  $c = \mathbf{0}$  implies  $\mathbf{x}^c - \mathbf{x}^{c'} = 0$  because for each binomial it is assumed that  $\mathbf{x}^c \succ \mathbf{x}^{c'}$ . The *Graver basis* for  $\mathcal{C}$  consists of all primitive binomials lying in the corresponding code ideal and is denoted by  $\text{Gr}_{\mathcal{C}}$ .

For a binary code  $\mathcal{C}$ , a codeword  $c$  in  $\mathcal{C}$  is a *circuit* if its support is minimal with respect to inclusion. Each binomial associated with such a codeword is also called a circuit. It follows that every binomial which is a circuit is also primitive.

In [15] circuits were defined by the additional condition that their entries are relatively prime. In the binary case this condition can be omitted. The extension to codes over an arbitrary prime field  $\mathbb{F}_p$ , however, cannot simply be accomplished by adding this condition. A different definition is required in order to obtain similar results for the code ideal over such a field (Prop. 3) since the exponents of the polynomials are treated as vectors in  $\mathbb{F}_p^n$ . The arising difficulties in this case are illustrated as follows.

**Example 1.** Consider a linear code  $\mathcal{C}$  over  $\mathbb{F}_7$  generated by

$$G = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 3 & 0 \end{pmatrix}$$

and the code ideal  $I_{\mathcal{C}}$  in  $\mathbb{Q}[a, b, c]$ . The codeword  $(1, 3, 0)$  and all its multiples  $(2, 6, 0)$ ,  $(3, 2, 0)$ ,  $(4, 5, 0)$ ,  $(5, 1, 0)$  and  $(6, 4, 0)$  have minimal support with respect to inclusion and except for  $(2, 6, 0)$  and  $(6, 4, 0)$  their entries are relatively prime. However, for the codeword  $(3, 2, 0) = (3, 0, 0) - (0, 5, 0)$ ,

the corresponding binomial  $a^3 - b^5$  is not primitive since  $a^2 - b$  lies in the code ideal,  $a^2$  divides  $a^3$ , and  $b$  divides  $b^5$ . It follows that minimal support and relative primeness of entries are not sufficient to ensure that such a binomial is primitive.

Another drawback is that different representations of a codeword yield binomials with distinct attributes. Writing  $(2, 6, 0) = (2, 0, 0) - (0, 1, 0)$  yields the primitive binomial  $a^2 - b$ . However, expanding  $(2, 6, 0) = (0, 6, 0) - (5, 0, 0)$  gives the binomial  $b^6 - a^5$ , which is not primitive because the binomial  $b^2 - a^4$  corresponding to  $(3, 2, 0) = (0, 2, 0) - (4, 0, 0)$  also belongs to the code ideal.  $\diamond$

Note that for toric ideals, when the exponents are viewed as vectors in  $\mathbb{Z}^n$ , the condition of relative primeness of the entries together with the minimality of the support is sufficient to guarantee that the corresponding binomial is primitive since for any  $c \in \mathbb{Z}^n$  the representation  $c = c^+ - c^-$  is unique. Motivated by the above example, a more general definition of circuits for codes over arbitrary prime fields is required: A binomial  $\mathbf{x}^c - \mathbf{x}^{c'}$  in  $I_{\mathcal{C}}$  with  $c' \neq \mathbf{0}$  is called a *circuit* if it is a primitive binomial and its support is minimal with respect to inclusion. In the non-binary case the attribute of being a circuit is tied to the binomial associated with a codeword rather than the codeword itself because a codeword may be associated with both, a primitive and a non-primitive binomial as the above example has demonstrated. With this more general definition, denote by  $C_{\mathcal{C}}$  all circuits lying in the code ideal  $I_{\mathcal{C}}$ . Finally, denote the universal Gröbner basis for the code ideal  $I_{\mathcal{C}}$  by  $\mathcal{U}_{\mathcal{C}}$ .

**Proposition 3.** For a linear code  $\mathcal{C}$  over  $\mathbb{F}_p$ ,  $C_{\mathcal{C}} \subseteq \mathcal{U}_{\mathcal{C}} \subseteq \text{Gr}_{\mathcal{C}}$ .

*Proof.* The inclusion  $\mathcal{U}_{\mathcal{C}} \subseteq \text{Gr}_{\mathcal{C}}$  is proved in [14]. So it remains to show that  $C_{\mathcal{C}} \subseteq \mathcal{U}_{\mathcal{C}}$ .

Let  $\mathbf{x}^{c'} - \mathbf{x}^{c''} \in I_{\mathcal{C}}$  be a circuit corresponding to the codeword  $c = c' - c'' \in \mathcal{C}$ . Put  $s = \deg(c')$  and  $t = \deg(c'')$  and choose an elimination order  $\succ$  such that  $\{x_i \mid i \notin \text{supp}(c)\} \succ \{x_i \mid i \in \text{supp}(c)\}$  and the monomials in  $\{x_i \mid i \in \text{supp}(c)\}$  are first compared by their  $\omega$ -degree, where  $\omega_i = t$  whenever  $i \in \text{supp}(c')$  and  $\omega_i = s$  whenever  $i \in \text{supp}(c'')$  and ties are broken by any lexicographic order with  $\{x_i \mid i \in \text{supp}(c')\} \succ \{x_i \mid i \in \text{supp}(c'')\}$ . For this order,  $\mathbf{x}^{c'} \succ \mathbf{x}^{c''}$  because of  $c' \cdot \omega = st = c'' \cdot \omega$  and the chosen tie breaker.

Claim that  $\mathbf{x}^{c'} - \mathbf{x}^{c''} \in \mathcal{G}_{\succ}(I_{\mathcal{C}})$ . Indeed, since this binomial belongs to  $I_{\mathcal{C}}$ , it must be reduced to zero by binomials in  $\mathcal{G}_{\succ}(I_{\mathcal{C}})$  and in particular, there must be a binomial  $\mathbf{x}^{v'} - \mathbf{x}^{v''} \in \mathcal{G}_{\succ}(I_{\mathcal{C}})$  such that  $v = v' - v'' \in \mathcal{C}$  and its leading term  $\mathbf{x}^{v'}$  divides  $\mathbf{x}^{c'}$ . But then  $\text{supp}(v') \subseteq \text{supp}(c')$  and by the choice of the monomial order it follows  $\text{supp}(v'') \subseteq \text{supp}(c)$ . Hence,  $\text{supp}(v) \subseteq \text{supp}(c)$ . But as  $c$  has minimal support this inclusion cannot be proper and so  $\alpha c = v$  for some  $\alpha \in \mathbb{F}_p$ .

Two cases occur. First, consider the case  $\alpha = 1$ , i.e.,  $c = v$ . Then by the choice of monomial order it can be further deduced that  $\mathbf{x}^{v'} = \mathbf{x}^{c'}$ . Otherwise,  $\mathbf{x}^{c'}$  would contain more variables than  $\mathbf{x}^{v'}$ , say  $\mathbf{x}^{c'} = \mathbf{x}^{v'} \mathbf{x}^w$  for some  $w \neq \mathbf{0}$  and then these missing variables must appear in the second term,  $\mathbf{x}^{v''} = \mathbf{x}^{c''} \mathbf{x}^{\mathbf{p}-w}$ , where  $\mathbf{p}$  is the all- $p$  vector. But then

$$v' \cdot \omega = (s - |w|)t < st + pt - |w|t = st + (p - |w|)t = v'' \cdot \omega$$

which contradicts  $\mathbf{x}^{v'} \succ \mathbf{x}^{v''}$ .

Second, consider the case  $\alpha \neq 1$ . Because of  $\text{supp}(v') \subseteq \text{supp}(c')$  either  $\text{supp}(v') \subsetneq \text{supp}(c')$  or  $\text{supp}(v') = \text{supp}(c')$ .

First assume that  $\text{supp}(v') = \text{supp}(c')$  and thus  $\text{supp}(v'') = \text{supp}(c'')$ . Since  $\mathbf{x}^{v'}$  divides  $\mathbf{x}^{c'}$  the monomial  $\mathbf{x}^{v''}$  cannot divide  $\mathbf{x}^{c''}$  for  $\mathbf{x}^{c'} - \mathbf{x}^{c''}$  is primitive. But as  $\text{supp}(v'') = \text{supp}(c'')$  the degree of  $\mathbf{x}^{v''}$  must be strictly greater than that of  $\mathbf{x}^{c''}$ . Hence,  $c'' \cdot \omega < v'' \cdot \omega$  because all  $x_i$  with  $i \in \text{supp}(c'')$  are weighted equally. Furthermore, as  $\text{supp}(v') = \text{supp}(c')$  and  $\mathbf{x}^{v'}$  divides  $\mathbf{x}^{c'}$  the same argument yields  $v' \cdot \omega < c' \cdot \omega$ . It follows that  $v' \cdot \omega < c' \cdot \omega = st = c'' \cdot \omega < v'' \cdot \omega$  contradicting the relation  $\mathbf{x}^{v'} \succ \mathbf{x}^{v''}$ .

Second assume that  $\text{supp}(v') \subsetneq \text{supp}(c')$ . Here the same inequality can be established when variables are shifted from  $\mathbf{x}^{v'}$  to  $\mathbf{x}^{v''}$  because all entries in  $\omega$  are positive. This will also lead to the contradiction that  $\mathbf{x}^{v'} \succ \mathbf{x}^{v''}$ .

In view of the two cases, it follows that  $\alpha = 1$ . By the first case this means that  $\mathbf{x}^{v'} = \mathbf{x}^{c'}$  and therefore  $\mathbf{x}^{c'} - \mathbf{x}^{c''} = \mathbf{x}^{v'} - \mathbf{x}^{v''}$ , as required.  $\square$

The proof justifies that a binomial of the form  $\mathbf{x}^c - 1$  is not considered as a circuit. Indeed, if  $\mathbf{x}^c - 1$  were the circuit considered in the proof, the weight vector  $\omega$  introduced there would be  $\omega = \mathbf{0}$  and so the contradiction  $v' \cdot \omega < v'' \cdot \omega$  could not be achieved.

This will be underpinned by the next example, which will show for a specific code that there exist primitive binomials  $\mathbf{x}^c - 1$  such that  $c$  has minimal support and does not belong to the universal Gröbner basis.

For binary codes, however, it can be shown that a binomial of the form  $\mathbf{x}^c - 1$  cannot belong to the universal Gröbner basis. To see this, assume that  $\mathbf{x}^c - 1$  lies in some reduced Gröbner basis for an arbitrary monomial order. Since the basis is reduced, the binomials of the form  $\mathbf{x}^{c'} - \mathbf{x}^{c''}$ , where  $c = c' - c''$  and  $c'' \neq \mathbf{0}$ , cannot belong to this Gröbner basis. But for any such binomial  $\mathbf{x}^{c'} - \mathbf{x}^{c''}$  there must be a binomial in the reduced Gröbner basis whose leading term divides the leading term of  $\mathbf{x}^{c'} - \mathbf{x}^{c''}$ . This binomial will then also divide the leading term of  $\mathbf{x}^c - 1$  contradicting the reducedness of the basis.

In view of binary linear codes, it will be shown that all three sets coincide. For non-binary linear codes, however, the next example will illustrate that the inclusions can be strict.



**Example 2.** In view of the code  $\mathcal{C}$  from Ex. 1, computations using the software package **Gfan** [9] exhibits that the set of circuits is

$$\begin{aligned} \mathcal{C}_{\mathcal{C}} = \{ & b - c^6, a - c^3, c^6 - b, b^2 - c^5, c^3 - a, a^2 - b, b - a^2, a^3 - c^2, \\ & c^2 - a^3, a^5 - c, c - a^5, c^5 - b^2, b^3 - c^4, c^4 - b^3, b^4 - c^3, \\ & b^4 - a, c^3 - b^4, b^5 - c^2, a - b^4, c^2 - b^5, b^6 - c, c - b^6 \} \end{aligned}$$

and that the universal Gröbner basis for the code ideal  $I_{\mathcal{C}}$  is indeed a proper superset,

$$\begin{aligned} \mathcal{U}_{\mathcal{C}} = \mathcal{C}_{\mathcal{C}} \cup \{ & bc - 1, a^2c - 1, ab^3 - 1, b^2 - ac^2, ab - c^2, b^3 - ac, ac^2 - b^2, \\ & c^2 - ab, ab^2 - c, ac - b^3, c - ab^2 \} \cup \{ a^7 - 1, b^7 - 1, c^7 - 1 \} \end{aligned}$$

Moreover, the universal Gröbner basis  $\mathcal{U}_{\mathcal{C}}$  properly lies inside the Graver basis  $\text{Gr}_{\mathcal{C}}$  since  $ac^4 - 1$  is a primitive binomial that belongs to  $\text{Gr}_{\mathcal{C}}$  but not to  $\mathcal{U}_{\mathcal{C}}$ . To see this, note that the binomial  $ac^4 - 1$  corresponds to the codeword  $(1, 0, 4) = (1, 0, 4) - (0, 0, 0)$  and only a binomial of the form  $a^s c^t - 1$  with either  $s < 1$  and  $t \leq 4$  or  $s \leq 1$  and  $t < 4$  could contradict its being primitive. But clearly no such codeword exists. Note additionally that the corresponding codeword  $(1, 0, 4)$  has minimal support.  $\diamond$

In the following, binary linear codes will only be considered.

**Theorem 4** ([14]). *For a binary linear code  $\mathcal{C}$  the set of circuits  $\mathcal{C}_{\mathcal{C}}$  equals the Graver basis  $\text{Gr}_{\mathcal{C}}$ .*

Combining Thm. 4 and Prop. 3 yields the following important result.

**Corollary 5.** *For a binary linear code  $\mathcal{C}$  the universal Gröbner basis  $\mathcal{U}_{\mathcal{C}}$  of the corresponding code ideal  $I_{\mathcal{C}}$  consists of all primitive binomials.*

In [15] the author has shown that the total degree of any primitive binomial in the toric ideal  $I_A$  is bounded by  $(n - d)(d + 1)D(A)$ , where  $d \times n$  is the size of the matrix  $A$  and  $D(A)$  is an integer number depending only on the entries in  $A$ . This result makes use of the estimate  $|u_i| \leq D(A)$ ,  $1 \leq i \leq n$ , for any circuit  $u = (u_1, \dots, u_n) \in \ker(A)$ . Furthermore, the author has conjectured that an even better estimation holds, namely that the total degree is bounded by  $(d + 1)D(A)$ . In the notation of binary linear codes the row size is  $d = n - k$  and any entry of a codeword is either 0 or 1 and thus  $D(A)$  can be chosen to be 1. Hence, the proven estimate becomes  $k(n - k + 1)$ . However, it will be shown that the bound  $n - k + 1$  conjectured by the author indeed holds. Note that this bound corresponds to the Singleton bound on the minimum distance of linear codes which is attained with equality by the maximum distance separable (MDS) codes like the Reed-Solomon codes and their extended versions. In the binary case, only trivial MDS codes exist [10, 16].

**Proposition 6.** Let  $\mathcal{C}$  be a binary  $[n, k]$  code. If  $\mathbf{x}^c - \mathbf{x}^{c'} \in I_{\mathcal{C}}$  is primitive, then  $\text{wt}(c - c') \leq n - k + 1$  and for any generator matrix  $G$  of the code  $\mathcal{C}$  the submatrix  $G_{\underline{n} \setminus \text{supp}(c - c')}$  has rank  $k - 1$ .

*Proof.* According to Thm. 4 the primitive binomials of the code ideal  $I_{\mathcal{C}}$  are exactly the circuits. The latter are given by sets of minimally dependent column vectors of a parity check matrix for the code. Any parity check matrix is of size  $(n - k) \times n$  and has rank  $n - k$ . Such a matrix has at most  $n - k$  linearly independent columns, which implies that the Hamming weight of a circuit is at most  $n - k + 1$ .

To show the second assertion, let  $c \in \mathcal{C}$  be a circuit. As  $c$  is a codeword there is a non-zero information word  $x \in \mathbb{F}_2^k$  with  $x \cdot G = c$  for any generator matrix  $G$ . But  $x \cdot G_{\underline{n} \setminus \text{supp}(c)} = \mathbf{0}$  and so the matrix  $G_{\underline{n} \setminus \text{supp}(c)}$  cannot have maximal rank  $k$ . Suppose the rank of  $G_{\underline{n} \setminus \text{supp}(c)}$  is smaller than  $k - 1$ . By the dimension formula for linear maps,

$$k = \dim \ker G_{\underline{n} \setminus \text{supp}(c)} + \dim \text{im } G_{\underline{n} \setminus \text{supp}(c)} < \dim \ker G_{\underline{n} \setminus \text{supp}(c)} + (k - 1)$$

and so  $\dim \ker G_{\underline{n} \setminus \text{supp}(c)} > 1$ . Thus there must be another information word  $x' \in \mathbb{F}_2^k$  with  $x' \cdot G_{\underline{n} \setminus \text{supp}(c)} = \mathbf{0}$ . Put  $c' = x' \cdot G$ . So for each index  $i$  in  $\underline{n} \setminus \text{supp}(c)$ ,  $c'_i = x' \cdot G_{\{i\}} = 0$  and thus  $\text{supp}(c') \subseteq \text{supp}(c)$ . But the encoding is one-to-one and so the codeword  $c'$  is distinct from  $c$ . It follows that  $\text{supp}(c') \subsetneq \text{supp}(c)$  contradicting the hypothesis that  $c$  is a circuit. Hence the rank of  $G_{\underline{n} \setminus \text{supp}(c)}$  must be equal to  $k - 1$ .  $\square$

The converse of this assertion also holds.

**Proposition 7.** Let  $\mathcal{C}$  be a binary  $[n, k]$  code with generator matrix  $G$ . Every binomial in the code ideal  $I_{\mathcal{C}}$  associated with a codeword  $c$  of Hamming weight less than or equal to  $n - k + 1$  and such that  $G_{\underline{n} \setminus \text{supp}(c)}$  has rank  $k - 1$  is primitive.

*Proof.* Consider a codeword  $c \in \mathcal{C}$  with Hamming weight  $\leq n - k + 1$  and such that  $G_{\underline{n} \setminus \text{supp}(c)}$  has rank  $k - 1$ . Then  $c$  has at least  $k - 1$  entries that are zero and so by hypothesis, among those one can find exactly  $k - 1$  coordinates  $J \subseteq \underline{n} \setminus \text{supp}(c)$  such that  $G_J$  has rank  $k - 1$ . But as the generator matrix  $G$  has rank  $k$  there must be another column in  $G$ , say indexed by  $i$ , with  $c_i = 1$ , and such that  $G_{J \cup \{i\}}$  is a  $k \times k$  matrix of rank  $k$ ; that is,  $J \cup \{i\}$  is an information set. By Cor. 2, the binomial  $x_i - \mathbf{x}^{c - e_i}$  belongs to the reduced Gröbner basis for a particular lexicographic order. Since every binomial in a reduced Gröbner basis for  $I_{\mathcal{C}}$  is primitive (see Prop. 3), the binomial  $x_i - \mathbf{x}^{c - e_i}$  is primitive. But if  $\mathbf{x}^u - \mathbf{x}^{u'}$  is primitive, then any other binomial  $\mathbf{x}^v - \mathbf{x}^{v'}$  with  $u - u' = v - v'$  is also primitive. Hence, each binomial associated with the codeword  $c$  is primitive, too.  $\square$

**Theorem 8.** *Let  $\mathcal{C}$  be a binary  $[n, k]$  code. The universal Gröbner basis for the corresponding code ideal  $I_{\mathcal{C}}$  is given by the set*

$$\mathcal{U}_{\mathcal{C}} = \{ \mathbf{x}^c - \mathbf{x}^{c'} \mid c - c' \in \mathcal{C}, \text{wt}(c - c') \leq n - k + 1, \\ \text{rk}(G_{\underline{n} \setminus \text{supp}(c - c')}) = k - 1 \} \cup \{ x_i^2 - 1 \mid 1 \leq i \leq n \}.$$

*In other words, the universal Gröbner basis for the code ideal consists of all binomials which correspond to codewords that satisfy the Singleton bound and a particular rank condition.*

*Proof.* Prop. 6 and 7 state that a binomial  $\mathbf{x}^c - \mathbf{x}^{c'} \in I_{\mathcal{C}}$  is primitive if and only if it corresponds to a codeword of Hamming weight  $\leq n - k + 1$  and such that the submatrix  $G_{\underline{n} \setminus \text{supp}(c - c')}$  of any generator matrix  $G$  has rank  $k - 1$ . The result follows by applying Cor. 5.  $\square$

This result gives rise to a new class of binary linear codes whose codewords which fulfill the Singleton bound also satisfy the rank condition. A binary linear code  $\mathcal{C}$  is called a *Singleton code* if each non-zero codeword  $c$  with Hamming weight  $\leq n - k + 1$  has the property that the submatrix  $G_{\underline{n} \setminus \text{supp}(c)}$  has rank  $k - 1$  for any generator matrix  $G$  for  $\mathcal{C}$ .

If  $\mathcal{C}$  is a Singleton code, then by Thm. 8 the corresponding universal Gröbner basis can be combinatorially constructed. For this, note that if  $c$  is a codeword with Hamming weight  $s$ , then there are  $2^s - 2$  binomials associated with  $c$ .

**Example 3.** The third binary Hamming code  $\mathcal{C}$  is a  $[7, 4]$  code with weight distribution  $1, 0, 0, 7, 7, 0, 0, 1$ . By inspection, this is a Singleton code, i.e., for any codeword  $c$  of Hamming weight  $\leq 4$  holds  $\text{rk}(G_{\underline{7} \setminus \text{supp}(c)}) \leq 3$  for any generator matrix  $G$  for  $\mathcal{C}$ . Thus the universal Gröbner basis  $\mathcal{U}_{\mathcal{C}}$  consists of all binomials which correspond to the codewords with Hamming weight of at most 4.

Computations using **Gfan** [9] exhibit that the universal Gröbner basis consists of 147 binomials given by seven binomials of the form  $x_i^2 - 1$ ,  $1 \leq i \leq 7$ ,  $42 = 7 \cdot (2^3 - 2)$  binomials corresponding to the seven codewords of Hamming weight 3, and  $98 = 7 \cdot (2^4 - 2)$  binomials associated with the seven codewords of Hamming weight 4.  $\diamond$

Singleton codes are the parity check codes, the MDS codes, the binary Golay code and its parity check extension, the Simplex codes, and the first order Reed-Muller code and its dual. On the other hand, not all Hamming codes are Singleton ones. We will provide more details in an upcoming paper.

As a final remark, the authors in [11] have introduced a method for computing the Graver basis for a linear code  $\mathcal{C}$  over  $\mathbb{Z}_p$ , where  $p \geq 2$  is an integer, which amounts to computing the Gröbner basis of a certain ideal. Since in

the binary case the Graver basis coincides with the universal Gröbner basis, this provides another method for computing the universal Gröbner basis for a binary linear code.

## References

- [1] W. Adams and P. Loustau. *An Introduction to Gröbner Bases*. American Mathematical Society, 1994.
- [2] T. Becker and V. Weispfenning. *Gröbner Bases – A Computational Approach to Commutative Algebra*. Springer, 1998.
- [3] A.M. Bigatti and L. Robbiano. Toric ideals. *Mathematica Contemporanea*, 21:1–25, 2001.
- [4] M. Borges-Quintana, M.A. Borges-Trenard, P. Fitzpatrick, and E. Martinez-Moro. Gröbner bases and combinatorics for binary codes. *AAECC*, 19(5):393–411, 2008.
- [5] P. Conti and C. Traverso. Buchberger algorithm and integer programming. In *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, volume 539 of *Lecture Notes in Computer Science*, pages 130–139. Springer Berlin / Heidelberg, 1991.
- [6] A.B. Cooper. Towards a new method of decoding algebraic codes using Gröbner bases. *Transactions 10th Army Conf. Appl. Math. Comp.*, 93:293–297, 1992.
- [7] D. Cox, J. Little, and D. O’Shea. *Ideals, Varieties, and Algorithms*. Springer, 1996.
- [8] Fulton. *Introduction to Toric Varieties*. Princeton University Press, 1993.
- [9] Anders N. Jensen. Gfan, a software system for Gröbner fans and tropical varieties. Available at <http://home.imf.au.dk/jensen/software/gfan/gfan.html>.
- [10] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North Holland, New York, 1977.
- [11] I. Marquez-Corbella and E. Martinez-Moro. Algebraic structure of the minimal support codewords set of some linear codes. *Advances in Mathematics of Communications*, 5:233–244, 2011.
- [12] T. Mora and L. Robbiano. The Gröbner Fan of an Ideal. *Journal of Symbolic Computation*, 6:183–208, 1988.

- [13] M. Saleemi and K.-H. Zimmermann. Gröbner bases for linear codes. *International Journal of Pure and Applied Mathematics*, 62:481–491, 2010.
- [14] M. Saleemi and K.-H. Zimmermann. Linear codes as binomial ideals. *International Journal of Pure and Applied Mathematics*, 61:147–156, 2010.
- [15] B. Sturmfels. *Gröbner Bases and Convex Polytopes*. American Mathematical Society, 1996.
- [16] J.H. van Lint. *Introduction to Coding Theory*. Springer, Berlin, 1999.
- [17] V. Weispfenning. Constructing universal Gröbner bases. In *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes (AAECC-5)*, volume 356 of *Lecture Notes in Computer Science*, pages 408–417. Springer, 1987.